Perceived criticism

Gerlsma, C.; van Duijn, M.A.J.; Hale III, W.W; van Hout, Wiljo

Published in:
Psychiatry Research

DOI:
10.1016/j.psychres.2008.09.007

Publication date:
2009

Psychiatry Research 170 (2009) 234–240

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Perceived criticism: Associations with perceiver characteristics and interpersonal behaviour

Coby Gerlsma [a,*], Marijtje A.J. van Duijn [b], William W. Hale III [c], Wiljo J.P.J. van Hout [a]

[a] Clinical and Developmental Psychology, University of Groningen, Groote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands
[b] Sociology/Interuniversity Center for Social Science Theory and Methodology, University of Groningen, The Netherlands
[c] Child and Adolescent Studies, Utrecht University, The Netherlands

ARTICLE INFO

Article history:
Received 4 December 2007
Received in revised form 19 August 2008
Accepted 30 September 2008

Keywords:
Expressed Emotion
Measurement
Construct validity
Relationships
Family

ABSTRACT

Perceived Criticism (PC) evolved in the context of Expressed Emotion (EE) research and, like EE, predicts the course of various psychiatric disorders. However, little is known about PC's validity. We examined (in Study 1) to what extent PC reflects the perceiver's current depressive and marital complaints, whether PC measures reciprocal criticism that characterizes dyads rather than individuals, and (in Study 2) whether PC reflects actual interactive behaviour. Both studies compared a single-item with a multi-item measure of PC. In Study 1, general community couples completed self-reports of PC, depressed mood, and marital dissatisfaction, and expressed their feelings in a brief EE interview (Five Minute Speech Samples). Multilevel analyses suggested that PC was associated with both partners' expressions of criticism, and the perceiver's depressive and marital complaints. In Study 2, general community couples completed self-reports of PC and participated in a videotaped problem solving interaction. Interactions were coded with the Kategorien system für partnerschaftliche Interaktion (KPI). PC was related to the partner's verbal and nonverbal expression of criticism and the perceiver's expression of nonverbal negative behaviour. Overall, the pattern of findings fits in well with theoretical formulations regarding EE, and supports the usefulness of PC measures as practical alternatives for the assessment of EE.

1. Introduction

Expressed Emotion (EE) research has shown that daily interactions of psychiatric patients with a relative or partner who expresses criticism, hostility, and/or emotional over involvement adversely affects the course of various psychiatric disorders, such as schizophrenia, depression, anxiety, posttraumatic stress disorder (PTSD) and eating disorders (reviews in Kavanagh, 1992; Butzlaff and Hooley, 1998; Wearden et al., 2000; Hooley, 2007). Moreover, when viewed as a measure of converse social support, its ramifications are even broader than psychiatric disorders alone (Franks et al., 1992; Wearden et al., 2000). The latter review documents findings on EE's predictive value for the course of medical conditions as well (e.g. asthma, diabetes, epilepsy).

1.1. Measurement of EE

The benchmark instrument for the assessment of EE is the Camberwell Family Interview (CFI; Vaughn and Leff, 1976), a semistructured interview with the patient's partner or relative. The interview generally lasts between 1 and 2 h, is recorded on audiotape, and is coded by trained raters. Both verbal and nonverbal behaviour characteristics are used to derive codes on five dimensions, i.e. Criticism, Hostility, Emotional OverInvolvement (EOI), Warmth and Positive Remarks. In terms of predictive power, Criticism is generally considered to be the key component of the EE concept. The Criticism index consists of a frequency count of the number of critical remarks the relative expresses about the patient during the interview.

Although it provides a wealth of information, the CFI has practical drawbacks. As Hooley and Parker (2006) review, CFI-training is lengthy, expensive and difficult to obtain; CFI-rating is time-consuming as the interview takes 1–2 h to administer and an additional 2–3 h to code. Apart from these limitations, administration of the CFI requires the cooperation of the patient's spouse or relative but they may not always be available or willing to participate (Bachmann et al., 2006). Moreover, the CFI is not suitable for repeated measurement. Therefore, it cannot be used to monitor changes in EE-level.

In view of the practical limitations of the CFI, several shortcut alternatives have been developed. Recent reviews by Van Humbeeck et al. (2002), Hooley and Parker (2006), and Renshaw (2008) highlight the self report measure labeled Perceived Criticism (PC) as a possibly recommendable supplement to or alternative for the CFI. This measure addresses the patient's point of view by simply asking patients 'How critical is your spouse/relative of you?' As Hooley and

* Corresponding author. Tel.: +31 50 3637604; fax +31 50 3637602.
E-mail address: J.Gerlsma@rug.nl (C. Gerlsma).

0165-1781/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.psychres.2008.09.007

Parker (2006) and Renshaw (2008) review, high PC ratings predict poorer course of symptoms in patients with unipolar depression, obsessive–compulsive disorder, agoraphobia with panic disorders, and in patients with substance abuse problems. However, the generally robust and substantial relationship of EE-Criticism with the course of schizophrenic symptomatology has not yet been replicated with PC as the predictor. Moreover, in a sample of patients with OCD and panic disorder, Steketee et al. (2007) found that higher PC predicted worse therapy outcomes only when patients were upset about the PC. Furthermore, Fogler et al. (2007) report converse findings with regard to therapy outcome in a sample of socially phobic patients. While PC did not predict a change in social phobia symptoms, lower levels of PC did predict increased drop out rates. One possible explanation for these apparently conflicting findings was proposed by Renshaw (2007), who demonstrated that PC-ratings have more predictive power when they concern partners or relatives one shares a home with (rather than significant others one does not live with). Another likely possibility is that the predictive power of PC varies across diagnoses.

1.2. Research questions

On account of its ease of administration and potential for prediction, PC ratings are increasingly used in clinical outcome studies. It is, however, 'far from clear exactly what ratings of PC actually measure' (Hooley and Parker, 2006, p. 393). Findings for convergent validity indicate that PC is moderately related to CFI rated EE-status ($r = 0.51$ in the original Hooley and Teasdale, 1989 study), but shows only modest associations with CFI-rated Criticism (i.e. Hooley and Teasdale, 1989, reported $r = 0.27$; Fogler et al., 2007, reported Spearman's rho = .25). Lack of empirical data with regard to the concurrent validity of PC ratings hampers a straightforward interpretation of predictive findings (Van Humbeeck et al., 2002; Renshaw, 2007, 2008), especially where predictive relationships appear to differ with diagnostic (sub)type. Moreover, as Renshaw (2008, p. 16) notes, "to target PC with individual or couples/family-based interventions, we must gain a greater understanding of the underlying construct being assessed by the PC measure [..]." Finally, PC is a single-item measure. Although extremely easy to administer, single-item measures are less than psychometrically ideal; they are less reliable and stable, and prone to misclassifications (Nunnally, 1970).

This study aims to address several aspects of the construct validity of PC to learn more about what responses on PC measures mean. In particular, we examined to what extent PC (1) is associated with characteristics of the perceiver, (2) is a measure of reciprocal criticism that characterizes dyads rather than individuals, (3) reflects actual interactive behaviour, and (4) to what extent single-item and multi-item ratings of PC differ in this respect. Study 1 examines the first two research questions; Study 2 examines, in an independent sample, the third question. Both studies are used to address the fourth research question.

2. Study 1

As an estimate of the level of EE in a dyadic relationship, PC should be a function of the number of critical comments (CC) expressed by one partner about the other. If critical comments are taken to reflect well-intended but possibly harmful efforts to change one's partner's behaviour (Hooley, 2007; McNab et al., 2007), one might expect characteristics of both relationship partners to affect the amount of criticism that is expressed within a particular dyad. Likewise, PC is likely to be associated with characteristics of the perceived as well as the perceiver.

2.1. PC and perceiver characteristics: depressive and marital complaints

One likely correlate of PC is the perceiver's current mood state. Individuals with depressive symptoms tend to elicit more negative responses in their social environment, including criticism and rejection, while the elevated levels of rejection sensitivity that often accompany depressed mood may lead to a stronger focus on the possibly critical and rejecting aspects in their partner's behaviour rather than his or her good intentions (e.g. Horowitz and Vitkus, 1986; Coyne et al., 1991). In a similar vein, individuals who evaluate their relationship as unsatisfactory may be prone to perceive more instances of criticism from their partner and give it more (negative) weight than individuals who are, on the whole, satisfied being with their partner (see also Kwon et al., 2006); conversely, individuals who perceive high levels of criticism in their partner are likely to report more marital dissatisfaction. Both depressive and marital complaints may therefore contribute to ratings of PC above and beyond the number of CC actually expressed by the partner.

2.2. PC and reciprocal criticism

Several EE-studies have shown that patients actively contribute to negative escalations in the daily interaction with their critical relative: they, too, attack and counter-attack (Hahlweg et al., 1989; Hahlweg, 2005). Therefore, EE-criticism is taken to reflect a habitual pattern of reciprocal criticism that characterizes certain dyads rather than certain individuals (Hooley and Parker, 2006; Hooley, 2007). If PC is a valid measure of EE, we may expect to find substantial interrelationships between partners, both in the amount of PC they report about each other and in the number of critical comments (CC) they actually express about each other.

2.3. Methods

2.3.1. Subjects and procedure

The dataset in this study is an as yet unpublished part of the data collected by Gerlsma and Hale (1997). To avoid undue influence of severe symptomatology and patient status, we selected couples from the general community rather than a clinical population. Participants in EE research are usually contacted and interviewed during a stressful period in their life (Hooley, 1998). To increase comparability in this respect, we selected couples who were expected to pass through a relatively stressful stage, i.e. parents with a first child between 6 and 18 months of age. We contacted forty couples through day-care centres and swimming pools. All couples were visited at their home address by two research assistants. After a brief interview, which included the Five-Minute Speech Sample, partners filled in questionnaire booklets in separate rooms, in the presence of a research assistant.

Forty (heterosexual) couples from the general community participated in this study. The mean age of the women was 31 years (standard deviation 4.7; range from 21 to 41 years); the mean age of the men was 33 years (S.D. 4.9, range 23–48). Thirty-two (80%) couples lived together as married couples, whereas seven (17.5%) lived together unmarried, and one (2.5%) couple 'lived-apart-together'.

2.3.2. Measures of perceived criticism, critical comments, depressive and marital complaints

Perceived criticism (PC) was assessed by means of the Criticism subscale of the Dutch version of the Level of Expressed Emotion scale (LEE; Cole and Kazarian, 1988; Gerlsma and Hale, 1997). This subscale (LEE-PC) consists of five items, i.e. four items that were adapted from the Criticism subscale of Hahlweg et al.'s (1995) Familien Fragenbogen, a self report questionnaire aimed to assess the two major EE-dimensions CC and EOI (Wiedemann et al., 2002) (My partner tries to change me, gets annoyed when I want something from him/her, usually agrees with me (reverse scoring), shows me (s)he cares for me (reverse scoring)), and one item that was adapted from Hooley and Teasdale's (1989) PC measure (My partner is critical of me). The latter item was also used as the single-item measure of PC (SIPC).

The response format of all five items is a 4-point Likert type scale anchored untrue–true; the total score of the LEE-PC ranges therefore from 5 to 20, and from 1 to 4 for the SIPC. Reliability of the LEE-PC in terms of internal consistency was Cronbach's alpha = 0.65. The failure to meet the recommended alpha >0.70 standard appeared to be due to the scale's length rather than lack of internal consistency (homogeneity (mean inter-item correlation corrected for autocorrelation) = 0.24). We performed separate analyses for the single-item PC measure (SIPC) and for the LEE-subscale PC measure (LEE-PC).

Critical comments (CC) was derived from Five-Minute Speech Samples (FMSS; Magaňa et al., 1986). The FMSS requires the individual to express his or her thoughts and feelings about the relative or partner and their relationship in a 5 min monologue. The FMSS-criteria for a rating of Criticism are derived from the CFI and state that critical comments should express distinct disapproval or dislike of the other's behaviour or personality, either verbally (e.g., 'I hate it when he..') or nonverbally (as inferred from

intonation and tempo). In this study, the speech samples were coded by the first author, who is a qualified CFI as well as FMSS rater and who was not involved in the actual data collection.

The amount of marital complaints was assessed by means of the dissatisfaction subscale of the Interpersonal Problem Solving Inventory (IPSI; Lange, 1983), a self report that consists of four items with a 5-point Likert type response format anchored 1 = satisfied to 5 = dissatisfied (the total score ranges therefore from 4 to 20; Cronbach alpha = 0.72 in this study). The amount of depressive complaints was assessed by means of the Beck Depression Inventory (BDI; Beck et al., 1961) (Cronbach alpha = 0.83 in this study).

2.3.3. Statistical analysis

Multilevel analysis was used to take into account the inherent interdependence of the dyadic data and to distinguish the different roles of partners in a couple relationship. The model is also known as the Actor Partner Independence Model (APIM, see also, Kenny et al., 2006). MLwiN 2.02 (Rasbash et al., 2005) was used for RIGLS estimation of the models.

First a so-called empty model was estimated to partition the total variance in individual (partner) variance and dyad (couple) variance. The dyad variance can also be thought of as expressing the reciprocity between partners because it is the covariance between partners' PC scores. The accompanying intra-class (couple) correlation, defined as the covariance divided by total variance, gives a standardized reciprocity estimate. Next we examined to what extent the data fitted the two models regarding the PC concept outlined in the Introduction section, i.e. PC's association with the perceiver's depressive and marital complaints, and the extent to which PC reflects reciprocal criticism.

2.4. Results

2.4.1. Preliminary analyses

The mean total score for LEE-PC was 8.60 (S.D. 2.45) and the mean score for SIPC was 1.74 (S.D. 0.88). Means and standard deviations for CC were 0.23 (S.D. 0.46), for depression 4.53 (S.D. 4.90), and for marital dissatisfaction 7.95 (S.D. 2.50). Paired *t*-tests reveal that, within dyads, the men perceived on average more criticism than their partners (9.00 vs. 8.19; $t = 2.28$, df = 39, $P = 0.028$ for LEE-PC; 1.93 vs. 1.55; $t = 2.73$, df = 39, $P = 0.009$ for SIPC), whereas the women reported higher depression scores than their partners (5.56 vs. 3.59; $t = 2.03$, df = 38, $P = 0.05$).

The partners-within dyads agreed considerably in their perception of criticism ($r = 0.53$, $P < 0.01$ for LEE-PC; $r = 0.43$, $P < 0.01$ for SIPC). Within-dyad associations between partners' scores were $r = 0.13$, n.s. for CC, $r = 0.17$, n.s. for depression, $r = 0.51$, $P < 0.01$ for marital dissatisfaction.

2.4.2. Multilevel analysis

Results of the multilevel analyses are presented in Table 1. One extreme observation was removed from the analysis (a man with a high score on LEE-PC), found to be influential on the parameter estimates in all models. Discarding this observation resulted in more conservatively estimated effects. Furthermore, nine other couples and two men with incomplete data on their own and their partner's scores on depression, marital satisfaction and/or critical comments were removed from the analyses. The scores on LEE-PC and SIPC were slightly but not significantly higher for the excluded respondents. Because only one woman was observed to make more than one critical comment, this variable was dichotomized, indicating whether or not critical comments were made. To enhance comparability between SIPC and LEE-PC we took the mean of LEE-PC as dependent variable. The fixed effects in the table represent the average effects of the perceiver's and the partner's characteristics on PC. They are tested with *t*-tests, with degrees of freedom equal to the number of observations (59) minus the number of fixed parameters in the models. The random effects are the two variance components of the APIM, pertaining to the individual and the dyad variance. The individual and dyadic variances are larger for SIPC than for LEE-PC, reflecting the lower reliability of the single item measure in comparison to the multi-item scale.

In the null or empty model, the standardized reciprocity of 0.54 for LEE-PC and 0.25 for SIPC show that partners-within-couples agree to a considerable extent in the amount of criticism they perceive from each other.

In Model 1, which examined the contribution of the perceiver's depressive and marital complaints to PC, the fixed effects indicate that LEE-PC scores increase as a function of the partner's CC (0.39) and the perceiver's own marital dissatisfaction (0.043) and depressed mood (0.021). For the SIPC scores only a significant and larger effect for the number of CC (0.73) was found; although the parameter estimates of the perceiver's depressive and marital complaints were of similar size as for LEE-PC, they did not reach significance due to the SIPC larger variances.

Comparison of the variance components in this model with those found in the (empty) Model 0 shows that the covariates added in Model 1 seem to explain dyadic variance in particular: The drop in dyad variance yields lower degrees of reciprocity (0.20 and 0.12) than that

Table 1
Estimates and standard errors (S.E.) for the empty model and the two conceptualized models of perceived criticism.

	Model 0 empty model		Model 1 'PC and perceiver's depressive and marital complaints'		Model 2 'PC and reciprocal criticism'	
	LEE-PC	Single-item PC	LEE-PC	Single-item PC	LEE-PC	Single-item PC
	Estimate (S.E)	Estimate (S.E)	Estimate (S.E.)	Estimate (S.E.)	Estimate (S.E.)	Estimate (S.E.)
Fixed effects						
Intercept	1.78 (0.07)	1.80 (0.13)	1.21 (0.18)	1.04 (0.38)	1.61 (0.08)	1.58 (0.14)
Perceiver's:						
Critical comments	–	–	–	–	0.19* (0.13)	0.01 (0.26)
Depressed mood	–	–	0.02* (0.01)	0.02 (0.03)	–	–
Marital dissatisfaction	–	–	0.04** (0.02)	0.06 (0.05)	–	–
Partner's:						
Critical comments	–	–	0.39*** (0.11)	0.73*** (0.25)	0.50*** (0.12)	0.85*** (0.25)
Random effects						
Dyad variance	0.12 (0.05)	0.19 (0.15)	0.03 (0.03	0.09 (0.12)	0.08 (0.04)	0.15 (0.13)
Individual variance	0.10 (0.03)	0.58 (0.15)	0.10 (003)	0.52 (0.14)	0.09 (0.03	0.50 (0.13)
Standardized reciprocity[1]	0.54	0.25	0.20	0.12	0.49	0.23
Deviance	67.5	149.1	40.9	133.0	52.1	137.8

[1]The standardized reciprocity estimate is derived as the covariance (dyad variance) divided by the total (dyad + individual) variance.

* $P < .10$ two-sided *t*-test, with df = 63 for model 1 and df = 58 for model 2.

** $P < .05$ two-sided *t*-test, with df = 63 for model 1 and df = 58 for model 2.

*** $P < .01$ two-sided *t*-test, with df = 63 for model 1 and df = 58 for model 2.

found in the empty model, indicating that the covariates added in Model 1 explain primarily differences between couples rather than individuals.

Model 2 examined PC as a measure of reciprocal criticism. The fixed effects for both own and partner's CC indicate that LEE-PC scores increase as a function of the amount of criticism *both* partners expressed about each other in their FMSS's, with more weight of the partner's critical comments. SIPC does not seem to reflect reciprocal criticism; the only significant contribution in this model was by the partner's CC. Although the covariates in Model 2 do seem to explain some (predominantly dyadic) variance, this model is somewhat weaker in explaining the variance: According to the Deviance estimates, Model 1 yields the best fit for both SIPC and LEE-PC.

2.5. Discussion Study 1

Both the SIPC and the LEE-PC were significantly associated with the number of CC expressed by the partner, a finding that supports the construct validity of both measures. Furthermore, the multi-item LEE-PC appeared to be associated with the amount of CC expressed by the perceiver, evidencing that it taps the reciprocal aspect of EE-criticism, and with the amount of depressive and marital complaints reported by the perceiver. The difference between the two instruments might well be due to the broader, more heterogeneous operationalization of PC by the LEE-PC, that seems to tap more than partner's criticism alone (including, for instance, an item on partners showing care for each other). On the other hand, the SIPC did show similar and equally large effect sizes for marital and depressive complaints, but these estimates did not reach significance due to the larger variances associated with SIPC's somewhat weaker reliability.

It should be noted that these findings may be affected by the present study's shortcomings in method and design. For one thing, and in line with the arguments raised above, the strength of the links found in this study was restricted by our operationalization of CC and SIPC. For practical reasons, our SIPC had a 4-point scale rather than the original 10-point response format used in the Hooley and Teasdale (1989) PC measure. The use of the FMSS rather than the CFI yielded a restricted CC-score. Hence, we may expect the association between SIPC and CC to be even stronger when psychometrically optimal instruments are used to assess each.

Furthermore, to examine the extent to which PC reflects perceiver characteristics, we analyzed the influence of depressive and marital complaints but we did not take into account other possibly contributing characteristics, such as the level of acute or chronic stress and neuroticism, which might well evoke a plaintive response style, or social desirability and defensiveness, which might lead to a somewhat idealized account of the spouse's behaviour.

Generalizability of our findings is furthermore restricted by the rather small sample size and the use of data provided by healthy community participants. No individuals or couples from a clinical population were included. The implications of this restriction will be discussed in the overall discussion, encompassing findings and limitations of Study 2.

3. Study 2

Assessment of EE-Criticism, either by means of interview or self report, assumes that the instrument's scores reflect how patients and their relatives actually interact with each other in their home environment. This assumption has not yet been tested in a naturalistic setting. In laboratory situations, however, empirical support was reported for the CFI (Hooley, 1986; Simoneau et al., 1999; Chambless et al., 2002) and the FMSS (Hahlweg et al., 1989). With regard to PC, empirical findings are sparse. In a sample of 22 clinical and 21 control couples, Chambless et al. (2002) found an $r = 0.52$ correlation between the Hooley and Teasdale (1989) single-item PC measure and the number of critical comments partners made in a problem solving dyadic interaction. This finding supports the convergent validity of the Single-item PC measure. In Study 2 we aim to replicate Chambless et al.'s finding (Model 1) and extend it by examining several additional aspects of the validity of PC.

3.1. PC and verbal and nonverbal features of critical interactive behaviour

According to CFI criteria, EE-ratings are based on both verbal and nonverbal characteristics of the speaker's speech sample (Vaughn and Leff, 1976). Therefore, a significant and substantial association between PC ratings and the partner's actual verbal *and* nonverbal critical behaviour during a dyadic interaction will strengthen the convergent validity of PC as a measure of EE (Model 2).

3.2. PC and reciprocal critical interactive behaviour

As was mentioned in the introduction to Study 1, EE is best interpreted as a measure of reciprocal criticism that characterizes dyads rather than individuals (e.g. Hooley and Parker, 2006). Ratings of PC may, therefore, be expected to relate to the actual critical behaviour of the perceived as well as the perceiver. Model 3 examines PC as a reflection of reciprocal verbal Criticism; Model 4 examines PC as a reflection of reciprocal verbal and nonverbal Criticism.

3.3. Methods

3.3.1. Subjects and procedure

The dataset used in this study is an as yet unpublished part of the data collected by Pielage (2006). Thirty-four couples from the general community responded to advertisements that invited both happy and distressed couples to participate in a study on stress and intimate relationships. The distressed ($N = 18$) and happy ($N = 15$) couples did not differ in marital status, education, relationship duration, or number of children. The mean duration of their relationships was 24.3 years (range 1–66 years; S.D. = 15.7); 82% of the couples had children (the number ranged between 0 and 5 with a mean of 1.92, S.D. = 1.27). Age of the participants ranged from 28 to 85 (mean age 50.14 years; S.D. = 11.04). Mean level of education was 4.44 (S.D. = 2.10) on a continuum from 1 (primary school only) to 8 (completed university education).

The data were collected at the Department of Clinical Psychology. After a brief introduction, the couples were asked to discuss a major relationship problem for 10 min and to try and reach a mutually satisfying solution. To familiarize themselves with the laboratory situation, the partners first practiced for a few minutes by discussing in front of the camera how they first met. Both the practice session and the problem solving interaction were in private, that is, without a research assistant present. After the video-interaction, the partners separately filled in the self reports on PC, in the presence of a research-assistant.

3.3.2. Measurement of PC and verbal and nonverbal interactive behaviour

PC was assessed in exactly the same way as in Study 1, i.e. by means of the LEE-PC and the SIPC .

Like most studies in this realm, we used the Kategoriensystem für partnerschaftliche Interaktion (KPI; Hahlweg et al., 1984) to code verbal and nonverbal behaviour in the problem solving interactions. The 10-min videotaped interaction is divided into coding units, i.e. verbal responses that are homogeneous in content without regard to its duration or syntactical structure. The unit ends when the speaker voluntarily stops or the partner successfully interrupts. For each unit (i.e. speaking turn), both partners are assigned a verbal and a nonverbal code. Criticism is 1 of the 12 categories of verbal behaviour in the KPI, and involves a specific critical comment or rejection of the partner. Ratings for nonverbal behaviour are based on facial and bodily cues as well as the tone of the voice.

Nine coders participated in the KPI-coding. Eight research students of our department were trained by a criterion coder (S.B. Pielage); both the criterion coder and the first author were trained by members of Professor Hahlweg's team in Braunschweig, Germany. All videotaped interactions were co-rated by two independent coders; disagreements were solved through discussion and consensus. Ten percent of the coded tapes were used for reliability checks with the criterion coder. The average percentage of agreement was 80% for the verbal and 75% for the nonverbal codes (Pielage, 2006).

3.4. Results

3.4.1. Preliminary analyses

The mean score for LEE-PC was 9.62 (S.D. 4.07) and the mean score for Single-item PC was 1.93 (S.D. 1.07) in this sample. Mean and

standard deviation for KPI–Criticism was 0.048 (S.D. 0.076); this score means that on average approximately 5% of a subject's speaking units were coded as critical comments. For KPI-negative nonverbal behaviour the mean score was 0.34 (S.D. 0.27). Within dyads, there were no significant ($P>0.05$) differences between men and women. Not surprisingly, the respondents in happy couples scored significantly lower on all these variables than respondents from distressed couples.

As in Study 1, the partners-within dyads agreed considerably in the perception of criticism ($r=0.67$, $P<0.01$ for LEE-PC; $r=0.49$, $P<0.01$ for SIPC). Within-dyad associations between partners' KPI scores were $r=0.15$ ($P>0.05$) for criticism, and $r=.69$ ($P<.01$) for nonverbal negative interactive behaviour. The correlations of SIPC and LEE-PC with KPI-coded criticism read $r=0.22$ (n.s.) and $r=0.40$ ($P<0.05$).

3.4.2. Multilevel analysis

Results of the multilevel analysis are summarized in Table 2. Due to incomplete data, six couples and two men were excluded from the analyses. These respondents, evenly distributed over the distressed and happy group, did not differ significantly from the included respondents on any of the (available) variables.

The empty model (Model 0) shows considerable agreement between partners-within-couples about the amount of criticism they perceive from each other (standardized reciprocity is 0.69 for LEE-PC and 0.48 for SIPC). In the Models 1 to 4, in which various covariates are entered, LEE-PC and SIPC show somewhat different patterns.

Scores for LEE-PC quite consistently relate to the partner's critical behaviour during the dyadic interaction (0.62 in Model 1, 0.55 in Model 2, 0.62 in Model 3, and 0.54 in Model 4). Furthermore, scores for LEE-PC reflect the partner's nonverbal interactive behaviour as well (0.87 in Model 2; 0.75 in Model 4). Unlike Study 1, however, the evidence for the model of LEE-PC as a measure of reciprocal critical interactive behaviour was not unequivocal: the estimates for the verbal expression of criticism by the PC-perceiver were not significant in Models 3 and 4, while the perceiver's own nonverbal negative behaviour did relate to his or her LEE-PC scores (0.64 in Model 4).

As in Study 1, the variances of the parameter estimates were larger for SIPC than for LEE-PC, which might be taken to reflect the somewhat weaker reliability of a single-item measure. However, unlike Study 1, this dataset also shows smaller effect sizes for SIPC than for LEE-PC. Support for the most narrow interpretation of PC, i.e. as a reflection of the partner's expression of explicit criticism, was found in the two models that include characteristics of the partner only. In Model 1 the influence of the partner's verbal expression of Criticism was significant (0.47); in Model 2 the partner's verbal expression of criticism (0.41) as well as his or her negative nonverbal behaviour (1.17) both contributed significantly. However, this convergence with the partner's critical behaviour is rendered nonsignificant when the perceiver's own nonverbal negative behaviour is taken into account (estimate 1.35 in Model 4), suggesting a mediation effect.

It should be noted that, according to the Deviance estimates, Model 4 yields the best fit for both LEE-PC and SIPC ratings.

3.5. Discussion Study 2

We did not quite replicate the large effect size ($r=0.52$) that Chambless et al. (2002) found for the association between PC and KPI-coded Criticism. With $r=0.22$ and $r=0.40$ our effect sizes were small to medium. In the multilevel analysis, both PC-measures related to the partner's actual critical behaviour during the problem solving interaction, both the verbal and the nonverbal features. As in Study 1, this finding supports the convergent validity of both measures. However, the best fit of the data was provided by the model that included the critical behaviour of the partner as well as the PC-perceiver, which suggests a fair degree of reciprocity in criticism. While this reciprocity was not expressed verbally, it did show in the nonverbal negative behaviour of the PC-perceiver. Taking account of the perceiver's own negative nonverbal behaviour appears to reduce the contribution of the partner's criticism. Interpreting this as a mediation effect would probably mean that the partner's critical behaviour affects the perceiver's nonverbal negative behaviour, which, in turn, is translated by the perceiver in (or is an epiphenomenon of) higher PC-ratings. This suggests that such PC-ratings are some kind of 'felt criticism', or, as Hooley and Teasdale (1989) put it, an index of 'how much criticism is getting through' (p. 234). This mediation pattern appears to be most clearcut in the case of the SIPC.

Table 2
Estimates and standard errors (S.E.) for the empty model and the two conceptualized models of perceived criticism.

	Model 0 empty model		Model 1 'PC and partner's verbal expression of criticism'		Model 2 'PC and partner's verbal expression of criticism and non-verbal negative behaviour		Model 3 'PC and reciprocal criticism'		Model 4 'PC and reciprocal negative interactive behaviour'	
	LEE-PC	Single-item PC	LEE-PC	Single-item PC	LEE-PC	Single-item PC	LEE-PC	Single-item PC	LEE-PC	Single-item PC
	Estimate (S.E)	Estimate (S.E)	Estimate (S.E.)	Estimate (S.E.)	Estimate (S.E.)	Estimate (S.E.)	Estimate (S.E.)	Estimate (S.E.)	Estimate (S.E.)	Estimate (S.E.)
Fixed effects										
Intercept	1.99 (0.15)	1.99 (0.18)	1.64 (0.18)	1.73 (0.24)	1.39 (0.20).	1.37 (0.27)	1.54 (0.21)	1.57 (0.28)	1.15 (0.24)	1.05 (0.30)
Perceiver's										
Verbal: – criticism							0.18 (0.21)	0.37 (0.32)	0.11 (0.21)	0.23 (0.31)
Nonverbal – negative									0.64* (0.39)	1.35* (0.59)
Partner's										
Verbal: – criticism			0.62*** (0.21)	0.47* (0.31)	0.55** (0.21)	0.41* (0.29)	0.62*** (0.21)	0.37 (0.32)	0.54** (0.20)	0.32 (0.31)
Nonverbal – Negative					0.87** (0.39)	1.17** (0.55)			0.75* (0.39)	0.54 (0.60)
Random effects										
Dyad variance	0.50 (0.17)	0.58 (0.26)	0.39 (0.14)	0.49 (0.24)	0.29 (0.12)	0.28 (0.21)	0.40 (0.14)	0.50 (0.24)	0.29 (0.11)	0.31 (0.20)
Individual variance	0.23 (.063)	0.63 (0.17)	0.22 (0.059)	0.66 (0.18)	0.23 (0.065)	0.74 (0.21)	0.22 (0.060)	0.65 (0.18)	0.22 (0.062)	0.65 (0.18)
Standardized reciprocity[1]	0.69	0.48	0.65	0.43	0.54	0.27	0.65	0.44	0.56	0.32
Deviance	118.5	155.7	110.2	153.4	105.6	149.4	109.5	152.1	102.0	142.8

[1]The standardized reciprocity estimate is derived as the covariance (dyad variance) divided by the total (dyad + individual) variance.
* $P<0.10$ two-sided t-test.
** $P<0.05$ two-sided t-test.
*** $P<0.01$ two-sided t-test.

Overall, the results in this dataset were less clear than in Study 1. A (partial) explanation might be the possibly somewhat restricted range of PC and KPI–Criticism ratings, due to sample characteristics. For one thing, the KPI-criteria for a rating of criticism are quite stringent and require an explicit and unambiguously critical statement. Since half of the couples in this sample claimed to be (very) happy in their current intimate relationship, such explicitly critical utterances were quite sparse in that half of the sample. Furthermore, the sample was rather small, which certainly restricted statistical power, as well as the means to explore possible explanations of puzzling findings, such as the influence of gender differences and interactions between the various parameters. Nevertheless, in combination with the findings in Study 1, the data raise interesting considerations that enhance our knowledge about the meaning of PC-responses, and might be useful to guide future studies.

4. Overall discussion

PC is increasingly used in studies on the course of psychological (and medical) complaints and in therapy outcome studies of various psychological disorders. It is an extremely practical measure, with proven predictive qualities. PC was recently recommended as a possible alternative for the more cumbersome CFI. However, although it predicts, we hardly know what PC exactly measures (Hooley and Parker, 2006; Renshaw, 2008).

The most consistent finding in these two studies on the construct validity of two PC-measures was that PC-scores do converge with 'objective' ratings of the amount of criticism that partners actually express, either during an interview or during a problem solving interaction with each other. In line with EE-conceptualizations, PC converged with both verbal and nonverbal features of criticism. Furthermore, partners showed considerable agreement in their perceptions of criticism and individuals who perceived more criticism in their partner's behaviour also tended to *express* more criticism themselves, either verbally or non-verbally. This result is in line with the interpretation of EE-Criticism as an index of a habitual pattern of reciprocal critical behaviour within dyads (Hahlweg, 2005; Hooley and Parker, 2006; Hooley, 2007). Above and beyond the amount of criticism expressed within the dyad, the amount of depressive and marital complaints, as experienced by the PC-perceiver, also proved to be relevant.

The multi-item measure of PC differed somewhat from the single-item measure. As was mentioned before, the multi-item LEE-PC includes an item that appears to go beyond criticism in particular, and seem to address relational quality in general (i.e. 'Shows me (s)he cares for me'), while three other items refer to partners disagreeing, wanting to change each other, or growing irritable. Hence, the bandwidth (Briggs and Cheek, 1986) of the LEE-PC is clearly broader than the SIPC, which might well explain the former's stronger link with marital dissatisfaction. The bandwidth of any single-item measure is, of course, minimal, and with this limitation in mind the SIPC appears to address a remarkably complex and meaningful set of behaviour (see also Renshaw, 2008). The drawback of measurement with a single item, i.e. it's relatively weak reliability, was however also apparent in our data, showing up in the relatively large variances of the estimates involved. It should be noted in this context, however, that we used a smaller response range than the single-item PC-scale recommended by Hooley and Parker (2006); their 10-point Likert-scale will probably yield more reliable results (see also Kwon et al., 2006).

With regard to these recommendations, it warrants mention that they are based on the data from two independent but rather small samples. This compromised our statistical power and precluded the possibility to explore the influence of, for instance, gender, differences and similarities between partners, and finer-grained interaction characteristics. Furthermore, both samples consisted of healthy volunteers from the general community, which may have yielded a somewhat restricted range of scores, and, consequently, smaller effect sizes. Apart from this statistical consequence, one may also question the validity of PC-measurement in non-clinical settings: is the criticism that two healthy partners perceive in each other (as in our study) the same construct as the criticism that a patient perceives in his or her healthy partner (as in most EE-studies)? Our findings appear to support a fair degree of similarity in terms of construct validity; it is likely that there are, however, substantial differences in the content and the impact of the criticism exchanged in clinical as opposed to healthy samples. Finally, the data in the second study were provided by two rather extreme groups, i.e. couples who claimed to be happy together, and couples who reported relational distress. We chose this composition of the sample in order to widen the score range within the total group. A liability of this approach is, however, that the two subgroups may have yielded different patterns of results. Because of the small size of both subgroups we were unable to sufficiently analyze and control for such differences.

We found considerable convergence between PC and two 'objective' measures of criticism in this study, which we interpret as support for the validity of the PC-measures. However, it might be well to take into consideration the conceptual differences between PC and objectively observed criticism in order to shed yet a different light on the question as to what responses to PC-measures mean. For instance, lack of convergence may also call into question the 'objectivity' of the coding system that is used to rate criticism. A graphic illustration of the point was recently published by Weismann et al. (2006), who found cultural differences in the congruence between PC and CFI-Criticism. For white and Latino patients with schizophrenia they found effect sizes of approximately $r = 0.50$, whereas for black patients the congruence read $r = -0.07$. The authors argue that current methods for the definition of criticism in the home environment, like the CFI and KPI, may not be equally meaningful and valid for people from all cultural and ethnic backgrounds. Renshaw (2008) notes still other conceptual differences between PC and 'objective' measures of criticism in the home-environment. One likely source of divergence may be the fact that the respondent has a more representative and naturalistic sample of the partner's behaviour to take into account in his or her PC-rating than the researcher, who has to rely on a necessarily limited sample of behaviour, collected in the laboratory or during an interview. From this perspective, respondents' subjective ratings of their partner's criticism may be more reliable and (ecologically) valid than 'objective' ratings. Another source of divergence may rise when respondents include both constructive and destructive criticism in their PC-ratings. None of the coding systems used to rate critical comments in EE-research (i.e. CFI, FMSS, KPI) has codes for constructive criticism; a rating of criticism is by definition a rating of negative interpersonal communication. Consequently, filtering out the contribution of constructive criticism from ratings of PC will presumably enhance PC's predictive power and convergence with conventional EE-indices (see also McNab et al., 2007).

What our findings and the considerations raised above seem to make clear is that, as an alternative measure of EE, PC-measures have one tremendous asset: they are extremely short and easy to administer, while resembling a true EE-index remarkably well (see also Hooley and Parker, 2006; Renshaw, 2008). This resemblance implies that PC-measures also share one great liability with other EE-indices, i.e. that they are merely summaries of a chain of interactions that is essentially an ongoing and dynamic process. To better understand the link between EE/PC and the course of psychological disorders and gain more insight in the factors that contribute to the development of high-EE attitudes, we may need to start observing those chains of interaction. While this may seem a daunting mission indeed, recent developments in the methodology of the dynamic systems approach (e.g. Granic and Hollenstein, 2003; Granic, 2005) might help to make the enterprise both feasible and rewarding.

References

Bachmann, S., Bottmer, C., Jacob, S., Schröder, J., 2006. Perceived criticism in schizophrenia: a comparison of instruments for the assessment of the patient's perspective and its relation to relatives' expressed emotion. Psychiatry Research 142, 167–175.

Beck, A.T., Erbaugh, J., Ward, C.H., Mock, J., Mendelson, M., 1961. An inventory for measuring depression. Archives of General Psychiatry 4, 53–63.

Briggs, S.R., Cheek, J.M., 1986. The role of factor analysis in the development and evaluation of personality scales. Journal of Personality 54, 106–148.

Butzlaff, R.L., Hooley, J.M., 1998. Expressed emotion and psychiatric relapse: a meta-analysis. Archives of General Psychiatry 55, 547–552.

Chambless, D.L., Fauerbach, J.A., Floyd, F.J., Wilson, K.A., Remen, A.L., Renneberg, B., 2002. Marital interaction of agoraphobic women: a controlled behavioral observation study. Journal of Abnormal Psychology 111, 502–512.

Cole, J.D., Kazarian, S.S., 1988. The level of expressed emotion scale: a new measure of expressed emotion. Journal of Clinical Psychology 44, 392–397.

Coyne, J.C., Burchill, S.A.L., Stiles, W.B., 1991. An interactional perspective on depression. In: Snyder, C.R., Forsyth, D.O. (Eds.), Handbook of Social and Clinical Psychology. Pergamon, New York, pp. 327–349.

Fogler, J.M., Thompson, M.C., Steketee, G., Hofmann, S.G., 2007. Influence of expressed emotion and perceived criticism on cognitive–behavioral therapy for social phobia. Behaviour Research and Therapy 45, 235–249.

Franks, P., Shields, C., Campbell, T., McDaniel, S., Harp, J., Botelho, R.J., 1992. Association of social relationships with depressive symptoms: treating an alternative to social support. Journal of Family Psychology 6, 49–59.

Gerlsma, C., Hale III, W.W., 1997. Predictive power and construct validity of the level of expressed emotion (LEE) scale; Depressed out-patients and couples from the general community. British Journal of Psychiatry 170, 520–525.

Granic, I., 2005. Timing is everything: developmental psychopathology from a dynamic systems perspective. Developmental Review 25, 386–407.

Granic, I., Hollenstein, T., 2003. Dynamic systems methods for models of developmental psychopathology. Development and Psychopathology 15, 641–669.

Hahlweg, K., 2005. The shaping of individuals' mental structures and dispositions by others. Interaction Studies 6, 131–144.

Hahlweg, K., Reisner, L., Kohli, G., Vollmer, M., Schindler, L., Revenstorf, D., 1984. Development and validity of a new system to analyze interpersonal communication: Kategoriensystem für partnerschaftliche Interaktion. In: Hahlweg, K., Jacobsen, N. (Eds.), Marital interaction: Analysis and modification. Guilford, New York.

Hahlweg, K., Doane, J.A., Goldstein, M.J., Nuechterlein, K.H., Magana, A.B., Mintz, J., Miklowitz, D.J., Snyder, K.S., 1989. Expressed emotion and patient–relative interaction in families of recent onset schizophrenics. Journal of Consulting and Clinical Psychology 57, 11–18.

Hahlweg, K., Dürr, H., Müller, U., 1995. Familienbetreuung schizophrener Patienten. Psychologie Verlag Union, Weinheim, pp. 54–59.

Hooley, J.M., 1986. Expressed emotion and depression: interactions between patients and high- versus low-expressed emotion spouses. Journal of Abnormal Psychology 95, 237–246.

Hooley, J.M., 1998. Expressed emotion and psychiatric illness: from empirical data to clinical practice. Behavior Therapy 29, 631–646.

Hooley, J.M., 2007. Expressed emotion and relapse of psychopathology. Annual Review of Clinical Psychology 3, 329–352.

Hooley, J.M., Parker, H., 2006. Measuring expressed emotion: an evaluation of the shortcuts. Journal of Family Psychology 20, 386–396.

Hooley, J.M., Teasdale, J.D., 1989. Predictors of relapse in unipolar depressives: expressed emotion, marital distress, and perceived criticism. Journal of Abnormal Psychology 98, 229–235.

Horowitz, L.M., Vitkus, J., 1986. The interpersonal basis of psychiatric symptoms. Clinical Psychology Review 6, 443–469.

Kenny, D.A., Kashy, D.A., Cook, W.L., 2006. Dyadic Data Analysis. Guilford Press, New York.

Kavanagh, D.J., 1992. Recent developments in expressed emotion and schizophrenia. British Journal of Psychiatry 160, 601–620.

Kwon, J., Lee, Y., Lee, M., Bifulco, A., 2006. Perceived criticism, marital interaction and relapse in unipolar depression: findings from a Korean sample. Clinical Psychology and Psychotherapy 13, 306–312.

Lange, A., 1983. Interactionele probleem oplossings vragenlijst (Interactional Problem Solving Inventory) Deventer. Van Loghum Slaterus, The Netherlands.

Magaña, A.B., Goldstein, M.J., Karno, M., Miklowitz, D.J., 1986. A brief method for assessing expressed emotion in relatives of psychiatric patients. Psychiatric Research 17, 203–212.

McNab, C., Haslam, N., Burnett, P., 2007. Expressed emotion, attributions, utility beliefs, and distress in parents of young people with first episode psychosis. Psychiatry Research 151, 97–106.

Nunnally Jr., J.C., 1970. Introduction to Psychological Measurement. McGraw-Hill, New York (etc.).

Pielage, S.B., 2006. Adult attachment and psychosocial functioning. Unpublished dissertation. University of Groningen, The Netherlands.

Rasbash, J., Browne, W., Healy, M., Cameron, B., Charlton, C., 2005. MLwiN Version 2.02. Multilevel Models Project, Institute of Education, University of London, London.

Renshaw, K.D., 2007. Perceived criticism only matters when it comes from those you live with. Journal of Clinical Psychology 63, 1171–1179.

Renshaw, K.D., 2008. The predictive, convergent, and discriminant validity of perceived criticism: a review. Clinical Psychology Review 28, 521–534.

Simoneau, T.L., Miklowitz, D.J., Saleem, R., 1999. Expressed emotion and interactional patterns in the families of bipolar patients. Journal of Abnormal Psychology 107, 497–507.

Steketee, G., Lam, J.N., Chambless, D.L., Rodebaugh, T.L., McCullouch, C.E., 2007. Effects of perceived criticism on anxiety and depression during behavioral treatment of anxiety disorders. Behaviour Research and Therapy 45, 11–19.

Van Humbeeck, G., Van Audenhove, Ch., De Hert, M., Pieters, G., Storms, G., 2002. Expressed emotion: a review of assessment instruments. Clinical Psychology Review 22, 321–341.

Vaughn, C., Leff, J., 1976. The measurement of expressed emotion in families of psychiatric patients. British Journal of Social and Clinical Psychology 15, 157–165.

Wearden, A.J., Tarrier, N., Barrowclough, C., Zastowny, T.R., Rahill, A.A., 2000. A review of expressed emotion research in health care. Clinical Psychology Review 20, 633–666.

Weismann, A.G., Rosales, G.A., Kymalainen, J.A., Armesto, J.C., 2006. Ethnicity, expressed emotion, and schizophrenia patients' perceptions of their family members' criticism. The Journal of Nervous and Mental Disease 194, 644–649.

Wiedemann, G., Rayki, O., Feinstein, E., Hahlweg, K., 2002. The family questionnaire: development and validation of a new self-report scale for assessing expressed emotion. Psychiatry Research 109, 265–280.